International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 4, May-June 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Credit Card Fraud Detection

Nikitha Pradeep[1], Dr. A Rengarajan[2]

[1]Student, Department of MCA, Jain University, Bangalore, Karnataka, India
[2]Professor, Department of MCA, Jain University, Bangalore, Karnataka, India

ABSTRACT

Credit card plays a very vital role in today's economy and the usage of credit cards has dramatically increased. Credit card has become one of the most common method of payment for both online and offline as well as for regular purchases of a common man. It is very necessary to distinguish fraudulent credit card transactions by the credit card organizations so their clients are not charged for the purchases that they didn't make. Despite the fact that using credit card gives huge benefits when used responsibly carefully and however significant credit and financial damages could be caused by fraudulent activities as well. Numerous methods have been proposed to stop these fraudulent activities. The project illustrates the model of a dataset to predict fraud transactions using machine learning. The model then detects if it is a fraudulent or a genuine transaction. The model also analyses and pre-processes the dataset along with deployment of multiple anomaly detection using algorithms such as Local forest outlier and Isolation forest.

KEYWORDS: *Credit card, Fraud detection, Isolation Forest, Local outlier factor, Support Vector Machine, anomaly detection*

How to cite this paper: Nikitha Pradeep | Dr. A Rengarajan "Credit Card Fraud Detection" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.426-429, URL: www.ijtsrd.com/papers/ijtsrd41289.pdf

IJTSRD41289

I. INTRODUCTION

Fraud can be defined as criminal deception which focuses on a monetary or personal gain[1], or which intends to harm or destroy another person while not essentially resulting to direct legal consequences. Credit Card Fraud can be defined as a case where an individual makes use of someone else's credit card for their private reasons while the owner and the card issuing authorities are not aware of the situation that the credit card is being utilized by another person.

Fraud detection systems can be utilized when the fraudsters excel the fraud interference systems and process a fraudulent transaction. Due to rise in the E-Commerce sector, there has been a rapid growth in the use of credit cards for online shopping as well as normal purchases which has increased the rates of frauds that are related to credit cards usage.

Fraud detection is conducted by monitoring the behaviour of different users and detected any undesirable changes from it. There has been a growing amount of monetary losses due to credit card and several papers reported vast amounts of losses in different countries [2-4].

Many anomaly detection techniques, supervised and unsupervised, are applied to find the fraud data involved in the transactions. The supervised techniques like SVM, Decision trees, KNN, logistic regression and others offer better results and can solve the issue of detecting fraud to an extent [5]. Yet, these methods need labelled data to create the classifier with fraudulent and non-fraudulent behaviours. In unsupervised technique the data does not have to be labelled. It is based on the fact that fraudulent behaviour will act very differently than normal. Decision trees, logistic regression, neural networks, Support Vector Machines, and nearest neighbour algorithms are some of most commonly used methods. The objective of this model is to detect if any of the latest transactions are fraudulent or not by using algorithms like isolation forest, Local Outlier Factor, Support Vector Machine which helps in reducing the number of false positives and detecting the highest number of fraud in the transactions by also checking the accuracy that each algorithm provides.

II. PROBLEM DEFINITION

Credit card frauds are of three types :traditional card related frauds, merchant related fraud and Internet frauds[6] .

Research has been done on many models and several techniques has been found to prevent and detect credit card frauds.

Some credit card fraud transaction datasets can be imbalanced. An accurate fraud detection system must be capable of detecting or identifying the fraudulent transaction accurately and should make the detection viable in real-time transactions. Fraud detection are of two types which are anomaly detection and misuse detection.

In Anomaly detection the normal transaction is trained and the fraudulent data is identified using various techniques. Contradictorily, in a misuse fraud detection labelled transactions are used as normal or fraudulent transactions.

So, the misuse detection system comes under supervised learning and anomaly detection under of unsupervised learning [7]. This model can be then used to detect if transaction is fraudulent or not.

III. WORKING METHODOLOGY

The dataset is imported and pre-processed. Exploratory data analysis is performed to the classes with respect to its frequency. The result shows the number of normal and fraud transactions that are present in the dataset. The data is analysed in terms of amount and time. The figure obtained shows that amount is small for fraud transactions but there a lot of fraud transactions in respect to time. A sample of data is taken to predict the number of fraudulent and normal transactions. Dependent and independent variables are created to apply the model. If the column name is not 'class' it is considered as independent feature and vice versa. Finally, the model prediction is done using isolation forest algorithm and local outlier factor. And the result predicts that these two algorithms outperforms SVM to separate outliers.

IV. WORKING FLOW

Import the dataset and perform the data pre-processing steps. Perform data analysis to find the number of class with respect to frequency. From the analysis performed, the number of normal transaction and the fraudulent ones are obtained where the normal transactions are high compared to the fraudulent ones. The value 1 depicts fraudulent transactions and 0 depicts normal transaction. The dataset is then checked for the fraudulent transactions in terms of amount and time. A sample of data is taken to determine how many are valid and fraud cases. Then the independent and dependent features are created to apply the model. X and Y are created where X is taken as dependent feature and Y is taken as independent feature. Model prediction is done using isolation forest algorithm and local outlier factor algorithm. The accuracy score and classification report is printed.

V. RESULT AND DISCUSSION

Fig 1 shows the number of classes with respect to frequency, It means that the normal transactions are more than 2,50,000 whereas the fraudulent transactions are very less in the dataset

Fig 2 shows the fraudulent data with respect to amount and the output shows that the amount is small for fraud transactions. Fig 3 shows transactions with respect to time and the output shows that there are a lot of transactions with respect to time.

Fig 4 shows the output of fraudulent data in the dataset. The isolation forest algorithm and Local Outlier Factor algorithm outperforms the support vector machine algorithm

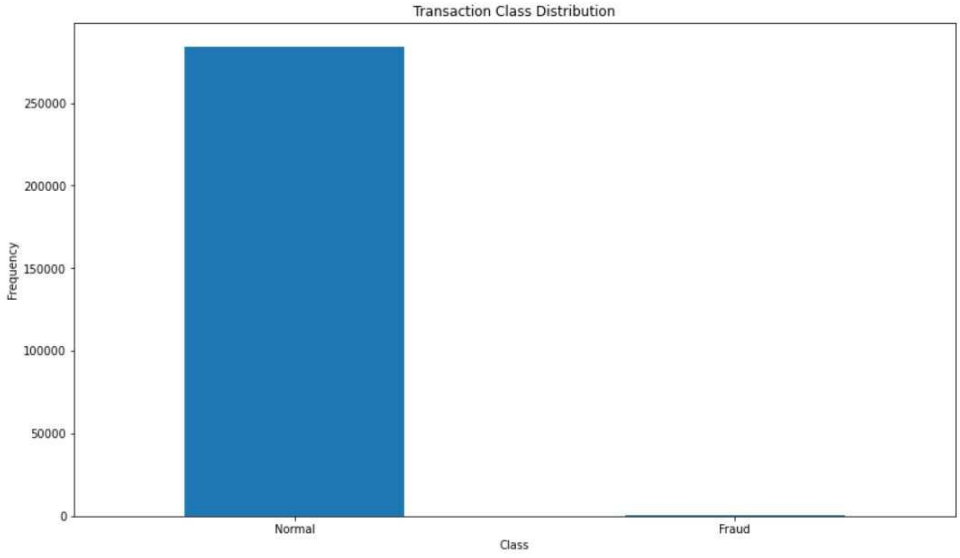

Fig 1

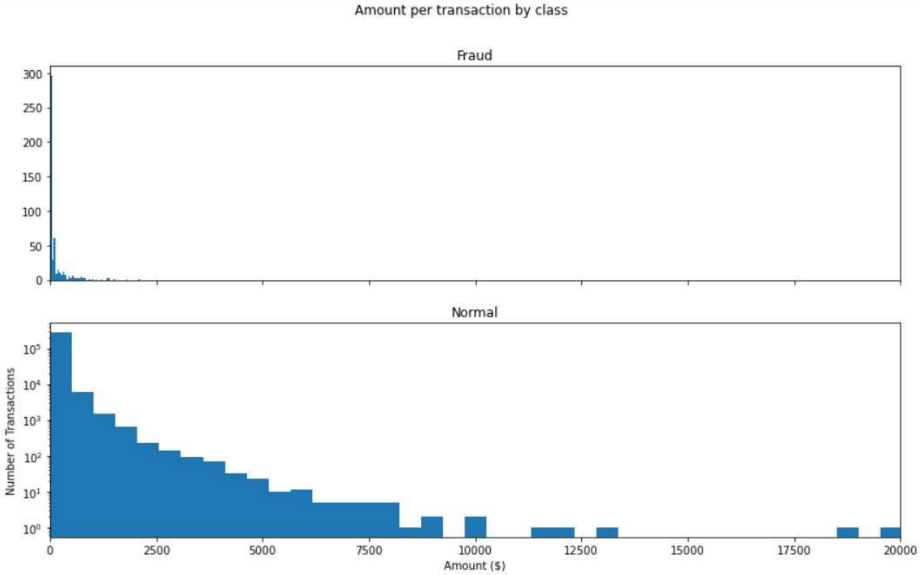

Fig 2

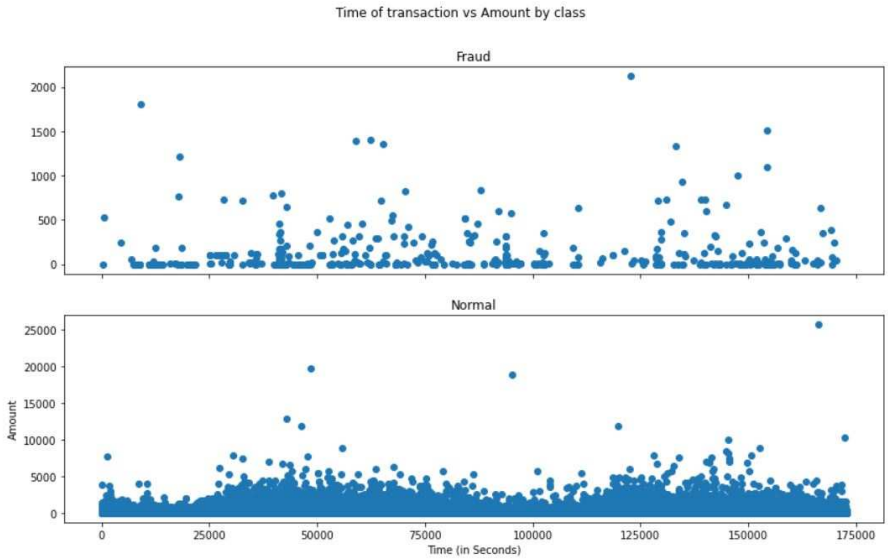

Fig 3

Isolation Forest: 73
Accuracy Score :
0.9974368877497279
Classification Report :

	precision	recall	f1-score	support
0	1.00	1.00	1.00	28432
1	0.26	0.27	0.26	49
accuracy			1.00	28481
macro avg	0.63	0.63	0.63	28481
weighted avg	1.00	1.00	1.00	28481

Local Outlier Factor: 97
Accuracy Score :
0.9965942207085425
Classification Report :

	precision	recall	f1-score	support
0	1.00	1.00	1.00	28432
1	0.02	0.02	0.02	49
accuracy			1.00	28481
macro avg	0.51	0.51	0.51	28481
weighted avg	1.00	1.00	1.00	28481

Support Vector Machine: 8516
Accuracy Score :
0.7009936448860644
Classification Report :

	precision	recall	f1-score	support
0	1.00	0.70	0.82	28432
1	0.00	0.37	0.00	49
accuracy			0.70	28481
macro avg	0.50	0.53	0.41	28481
weighted avg	1.00	0.70	0.82	28481

Fig 4: OUTPUT

73 errors has been detected using the isolation forest algorithm and it has an accuracy of 99.74% . Local Outlier Factor has detected 97 errors with 99.65% accuracy, whereas Support vector Machine detected 8516 errors accuracy with 70.09%. When comparing error precision & recall for 3 models, the Isolation Forest performed much better than the Local Outlier Factor and the detection of fraud cases is around 27 % Local Outlier Factor detection rate is 2 % and Support Vector Machine is 0%. So overall Isolation Forest Method performed much better in determining the fraud cases which is around 30%.

VI. CONCLUSION

The In this paper an analysis of credit card fraud identification was described on a publicly available dataset utilizing Machine Learning techniques such as Isolation Forest algorithm and Local Outlier Factor. The result has shown that the isolated forest is very efficient and outperforms in detecting anomalies in the case of the credit card. The use of this algorithm in credit card fraud detection system results in detecting or predicting the fraud probably in a very short span of time after the transactions has been made. This will eventually prevent the banks and customers from great losses and also will reduce risks.

ACKNOWLEDGMENT

I should convey my real tendency and obligation to Dr.DineshNilkhantDirector School of CS and IT, Dr. M N Nachappa DeanSchool of CS and IT, Program Coordinator: Dr.Bhuvana J,Project Coordinators: Dr. Lakshmi J V N and Dr.Gangothri R, Prof: Rengarajan A for their effective steerage and consistent inspirations all through my assessment work. Their ideal bearing, absolute coaction and second discernment have made my work gain.

REFERENCES

[1] A. Hobson, *The Oxford dictionary of difficult words*. Oxford University Press, USA, 2004.

[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical science,* vol. 17, no. 3, pp. 235-255, 2002.

[3] K. J. Leonard, "Detecting credit card fraud using expert systems," *Computers & industrial engineering,* vol. 25, no. 1-4, pp. 103-106, 1993.

[4] S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neural-network," in *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*, 1994, vol. 3: IEEE, pp. 621-630.

[5] A. R. Flarence, S. Bethu, V. Sowmya, K. Anusha, and B. S. Babu, "Importance of supervised learning in prediction analysis," *Periodicals of Engineering and Natural Sciences,* vol. 6, no. 1, pp. 201-214, 2018.

[6] L. Seyedhossein and M. R. Hashemi, "Mining information from credit card time series for timelier fraud detection," in *2010 5th International Symposium on Telecommunications*, 2010: IEEE, pp. 619-624.

[7] M. Smith, "The federal cyber role: How federal cybersecurity policy has affected the public and private sector", 2017, Utica College.